

*-Final Report of the Project:*

**Sub-mesoscale Ocean State Estimation in Shelf/Slope Regions Using  
Variational Data Assimilation to Integrate Ocean Observing System  
Data: Application to SW06**

Javier Zavala-Garay

Institute of Marine and Coastal Sciences  
Rutgers, The State University of New Jersey  
New Brunswick, NJ 08901-8521

phone: (732) 932-3433 fax: (732) 932-8578 email: [jzavala@marine.rutgers.edu](mailto:jzavala@marine.rutgers.edu)

CO-PI(s) name John L. Wilkin

Institute of Marine and Coastal Sciences  
Rutgers, The State University of New Jersey  
New Brunswick, NJ 08901-8521

phone: (732) 932-6555x251 fax: (732) 932-8578 email: [wilkin@marine.rutgers.edu](mailto:wilkin@marine.rutgers.edu)

Award Number: N000140810686

<http://www.myroms.org/espresso/>

**LONG-TERM GOALS AND RELEVANCE OF THIS PROJECT**

The long term goal is to develop a robust ocean prediction system for the North East US coast that efficiently exploits all the available observational platforms combined with the Regional Modeling System (ROMS). The system is now operating in real-time and forecasts of the state of the ocean six days in the future are posted daily at [www.myroms.org/espresso](http://www.myroms.org/espresso). The project summarized here supported: (i) development of the prognostic part of the forecast system (the forward model), (ii) implementation of variational data-assimilation of several remotely sensed and in-situ data streams, (iii) validation with hydrographic data from 2006 and 2007, notably that acquired during the ONR Shallow Water Acoustics 2006 (SW06) field program, and (iv) development of a prototype ensemble prediction system that will allow us to assign error-bars to the real-time deterministic forecast.

**OBJECTIVES**

To develop a prototype of an operational analysis and dynamical forecast system for mesoscale and sub-mesoscale variability in a coastal transition zone: The Mid-Atlantic Bight (MAB). The system should use advanced data assimilation and adjoint techniques to integrate a high-resolution 3-dimensional coastal model (The Regional Ocean Modeling System; ROMS) with data from a coastal observing system comprised of remotely-sensed Sea Surface Height Anomaly (SSHA) and Sea Surface Temperature (SST), surface current from high frequency radar (CODAR) installations, autonomous gliders, moorings and XBT/CTD acquired during the ONR Shallow Water Acoustics 2006 (SW06) field program.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>					
<b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.</b>					
1. REPORT DATE (DD-MM-YYYY) 4/15/11	2. REPORT TYPE FINAL	3. DATES COVERED (From - To) 01-APR-08 to 31-DEC-10			
<b>4. TITLE AND SUBTITLE</b> Sub-mesoscale Ocean State Estimation in Shelf/Slope Regions Using Variational Data Assimilation to Integrate Ocean Observing System Data: Application to SW06			5a. CONTRACT NUMBER N00014-08-1-0686		
			5b. GRANT NUMBER N/A		
			5c. PROGRAM ELEMENT NUMBER N/A		
<b>6. AUTHOR(S)</b> Zavala-Garay, Javier Wilkin, John, L.			5d. PROJECT NUMBER 10PR05047-00		
			5e. TASK NUMBER N/A		
			5f. WORK UNIT NUMBER N/A		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> RUTGERS THE STATE UNIVERSITY OF NEW JERSEY OFFICE OF RESEARCH AND SPONSORED PROGRAMS 3 RUTGERS PLAZA, ASB III NEW BRUNSWICK, NJ 08901-8559			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b> N/A		
<b>9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> Office of Naval Research Code 322 875 North Randolph Street Arlington, VA 22203-1995			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>  <b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION/AVAILABILITY STATEMENT</b> Approved for Public Release; distribution is Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> The long term goal is to develop a robust ocean prediction system for the North East US coast that efficiently exploits all the available observational platforms combined with the Regional Modeling System (ROMS). The system is now operating in real-time and forecasts of the state of the ocean six days in the future are posted daily at <a href="http://www.myroms.org/espresso">www.myroms.org/espresso</a> . The project summarized here supported: (i) development of the prognostic part of the forecast system (the forward model), (ii) implementation of variational data-assimilation of several remotely sensed and in-situ data streams, (iii) validation with hydrographic data from 2006 and 2007, notably that acquired during the ONR's Shallow Water Acoustics 2006 (SW06) field program, and (iv) development of a prototype ensemble prediction system that will allow us to assign error-bars to the real-time deterministic forecast.					
<b>15. SUBJECT TERMS</b> Ocean Modelling, Data Assimilation, Real-time Forecast Ocean Circulation					
<b>16. SECURITY CLASSIFICATION OF:</b> a. REPORT    b. ABSTRACT    c. THIS PAGE U            U		<b>17. LIMITATION OF ABSTRACT</b> UU	<b>18. NUMBER OF PAGES</b> 11	19a. NAME OF RESPONSIBLE PERSON Javier Zavala-Garay 19b. TELEPHONE NUMBER (Include area code) (732) 932-3433	

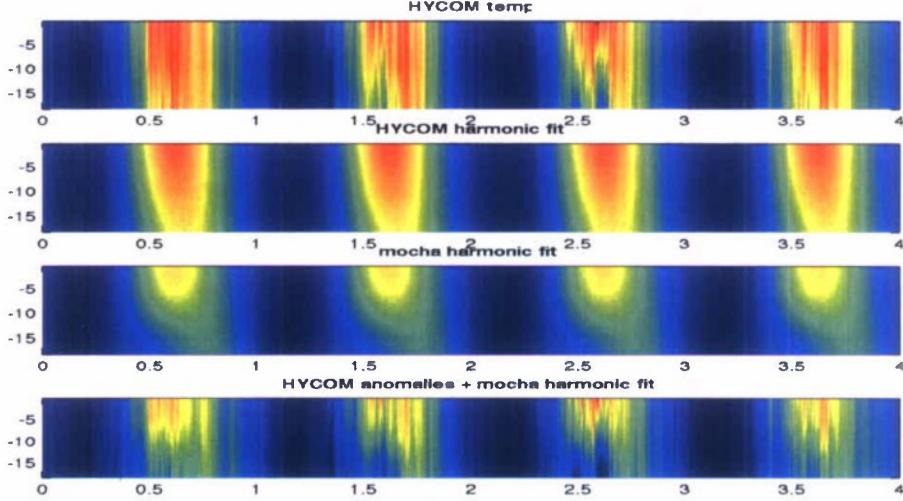
## APPROACH

The MAB continental shelf break is characterized by large horizontal and vertical gradients in water properties associated with the shelf break front, a feature susceptible to non-linear instabilities and strong interactions with Gulf Stream warm-core rings that impinge onto the continental slope (Ryan et al., 2001). The broad spectrum of forcing mechanisms that influence the shelf-break transition zone linking the circulation in the MAB and adjacent deep ocean make the region a challenging laboratory for testing coastal ocean model skill. Therefore the existence of the extensive archive of data acquired during the SW06 field campaign provides an excellent opportunity to test and validate the new data assimilation capabilities of ROMS in a transition area dominated by two different circulation regimes, and to prototype an adjoint-based operational ensemble system. To this end, two principal research tasks were developed in this project: (i) implementation of a data-assimilative analysis system in a region having two differing circulation regimes (the coastal zone and adjacent deep ocean), and (ii) implementation of an efficient yet rigorous ensemble forecast system that exploits the known statistical properties of the atmospheric weather or stochastic forcing of the region. These two topics are related through their use of a common set of modeling tools; namely, the tangent linear and adjoint versions of the nonlinear operator of ROMS. In the case of data assimilation, these codes underpin the formulation of strong-constraint 4-dimensional variational methods that allow the incorporation of any data that are a function of the model state variables. In the case of ensemble forecast system, a variant of the adjoint sensitivity driver is used to estimate the possible divergence of model forecasts due to the action of atmospheric weather, which highly affects the coastal zone and, due to its inherent chaotic nature, in an operational framework is not known beyond a few days.

## WORK COMPLETED (RESULTS)

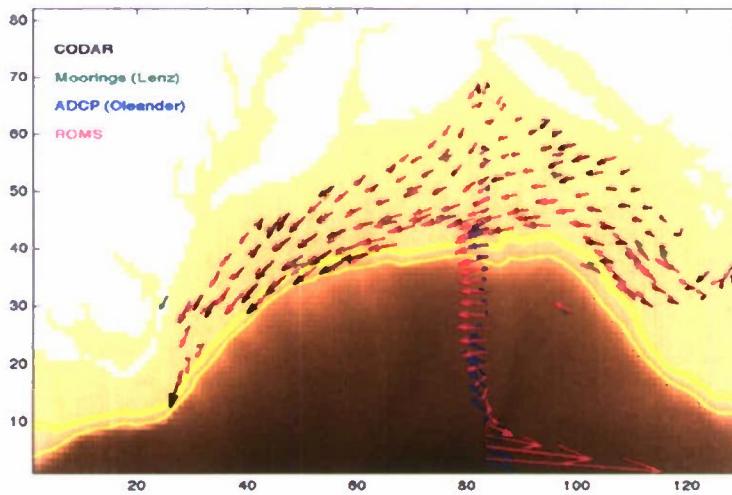
**Task 1: Limited area ROMS configuration.** A limited area ROMS model encompassing the MAB and adjacent deep ocean region was configured by J. Wilkin during year 1 of this project. The ESPRESSO (Experimental System for Predicting Shelf and Slope Optics) prototype system has a 5-km horizontal resolution, and 36-sigma levels. Atmospheric forcing data are taken from the North American Mesoscale (NAM) forecast system operated by NCEP. NAM is an implementation of WRF-NMM for the North American continent and adjacent seas, with a horizontal resolution of approximately 15 km. ROMS uses NAM surface air temperature, pressure and relative humidity, 10-m vector winds, rain, downward longwave radiation, and net shortwave radiation, to specify surface fluxes of momentum and buoyancy using standard bulk formulas. Real-time river runoff is prescribed from USGS river gauges. An early version of this model (reported in previous progress reports) used open boundary conditions taken from another ROMS application developed by Dr. Ruoying He that considers the MAB and Gulf of Maine region, and includes data-assimilative HYCOM at the boundaries and nudging to observed SST. Careful analysis of this model output revealed

that strong biases were introduced by the boundary forcing used and that this bias was also present in the HYCOM model. We therefore developed an improved version in which we replace the temperature and salinity seasonal cycle of HYCOM by that estimated from a large collection of hydrographic data. Figure 1 shows an example of such improved boundary forcing.



*Figure 1: Bias corrected temperature boundary forcing at a grid point in the north part of the Espresso domain. Harmonic analysis was performed in the HYCOM boundary forcing and the HYCOM seasonal cycle was replaced by the seasonal cycle estimated from observed historical data.*

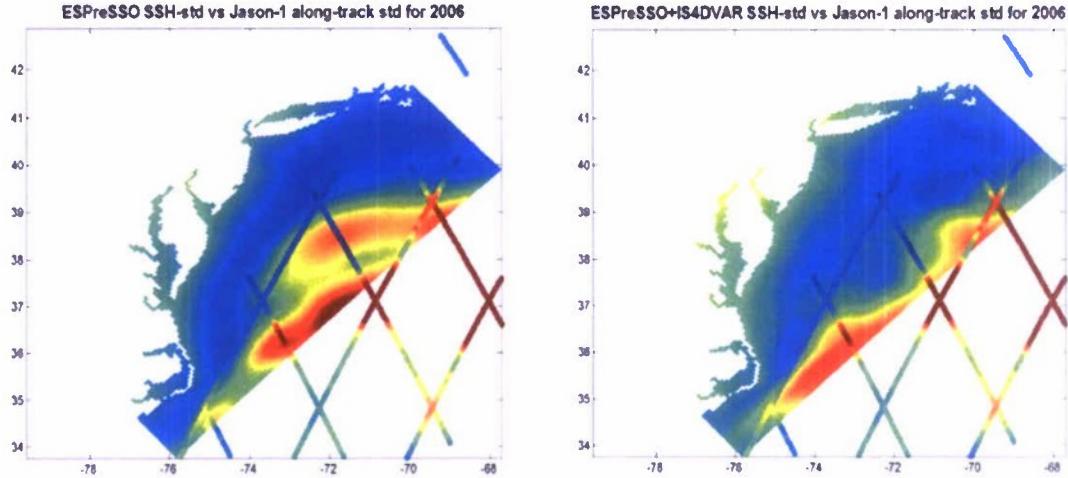
This seasonally bias-corrected boundary forcing from HYCOM highly improved the forward model and is the one used in the real time forecast system. Figure 2 shows a comparison between the mean surface currents with those observed by the HF CODAR radar, the ADCP onboard of the *Olcander*, and in-situ current meters.



*Figure 2: Comparison of Espresso mean surface currents with those measured by different observational platforms.*

The mean transports in the shelf are correctly simulated by ROMS while strong discrepancies occur near the boundaries where the ESPRESSO velocities are nudged to HYCOM.

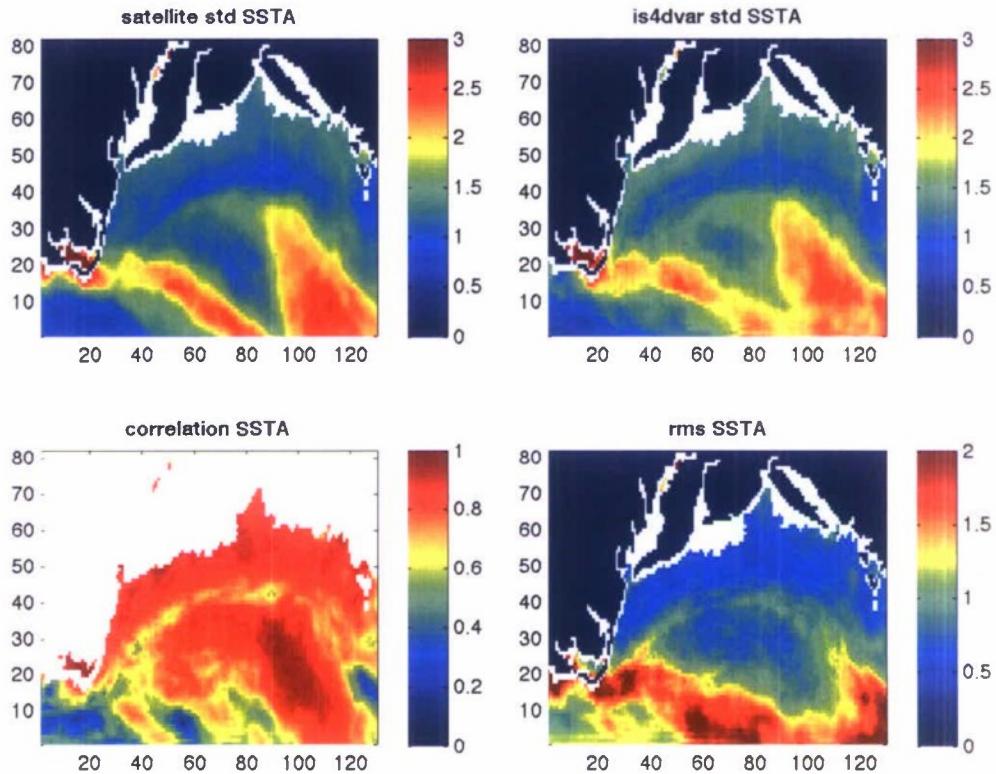
This new ESPRESSO application produces a reasonable seasonal cycle but fails in producing adequate mesoscale variability. This fact is illustrated in Fig 3 (left panel) that compares the SSH standard deviation of the model with that of the altimeter data.



**Figure 3: Standard deviation of the ESPRESSO application without (left) and with (right) data assimilation of altimeter data for 2006.**

**Task 2: Development of an operational data assimilation system.** We have prototyped an Incremental Strong-constraint 4-Dimensional Variational (IS4DVAR) data assimilation system for the ESPRESSO application. This task required selection of appropriate parameters such as appropriate decorrelation scales used in the IS4DVAR algorithm, number of iterations needed to converge to the minimum of the model misfit, refinement of the weighting of background and observation contributions to the cost function misfit, length of the assimilation window, etc. It also required considerable work in the collection and quality control of the observational products to be assimilated. The skill of the deterministic forecast (a single integration of the nonlinear model after data assimilation) was quantified for years 2006 and 2007 where extensive data has been collected by other groups and was made available to us.

The system is now assimilating satellite SST and altimeter data, hydrographic data from gliders, and surface currents from the existing high frequency radar (CODAR) array. As an example of the skill in nowcasting the surface mesoscale variability Figure 4 shows the standard deviation, correlation, and RMS error of daily SST anomalies during the 2 years period (2006 and 2007).



**Figure 4: Comparison between standard deviation of observed daily SST anomalies (top left) and ESPreSSO (top right) during 2006 and 2007. The solution in the interior (away from the boundaries, notably in the shelf and shelf-break) exhibits high skill (high correlation and low RMS error).**

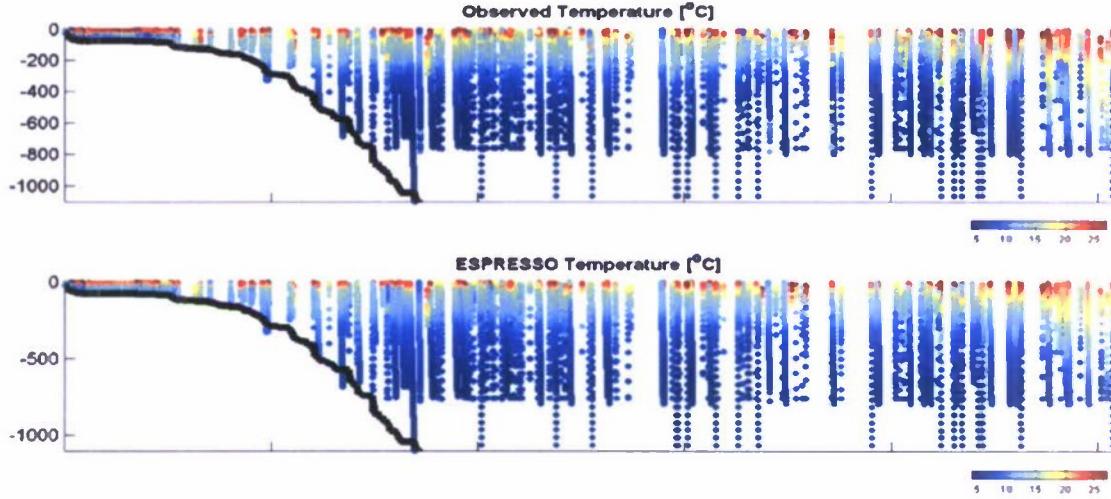
This project a lot of progress was made by including all these data streams into the assimilation system, but perhaps the one that we learned the most was the assimilation of along track data. It is a common practice to assimilate gridded SSH products into ocean models (Zavala-Garay et al, 2010). This is a good approximation in regions where mesoscale eddies have a long relatively isotropic length scales of the order of hundreds of km (Wilkin et al, 2002). SSH in coastal regions such as the MAB exhibits length and time scales that are shorter and more strongly anisotropic than in the deep ocean due to flow-bathymetry interactions. This presents challenges to statistically-based methods (e.g. optimal interpolation) typically used to produce spatial gridded maps from along-track data (Ducet et al., 200). Our approach therefore was to use variational methods to assimilate the along-track data into the ESPreSSO domain instead of the gridded products. The decision was also motivated by the fact that for the MAB there are strong differences between the along-track data and the AVISO gridded product. This difference is due partly to smoothing in time by the optimal interpolation, but principally because of the anisotropy of variability where Gulf Stream rings interact with the bathymetry of the continental shelf. A number of unanticipated methodologies were needed in order to implement the assimilation of along-track data in the ESPreSSO application, which are described next.

**Dynamic topography:** ROMS assimilates total SSH which is the sum of the mean SSH (or dynamic topography) plus the SSH anomaly due to mesoscale activity plus the tides. It is a common practice to use long term mean of the integration of the model as a proxy for the dynamic topography. We found however that even small biases in the model vertical structure of temperature and salinity will be noticeable in the mean SSH. More important, assimilation of SSH in a biased profile results in an incorrect vertical projection of the SSH. Therefore the Mean Dynamic Topography (MDT) was computed by IS4DVAR analysis of a regional high-resolution (4 km) 3-dimensional temperature and salinity climatology computed from historical hydrographic data by a member of our group. This MDT and the climatology itself (with appropriate error bars) was included in the assimilation experiments.

**Tides:** The variational data-assimilation method essentially consist in minimizing a cost function that measures the weighted sum of model-data misfit. Variations in the time dimension during the analysis interval is fully accounted for by sampling the model at the precise time and location that the observations were made. Any mismatches at high frequency in the modeled and observed SSHA therefore enter into the cost function. This is potentially problematic in coastal waters where the oceanic response of sea level to tidal and other high frequency forcing is substantially greater than in regions where sub-tidal frequency mesoscale variability dominates the dynamics. Anticipating the possibility that even small phase errors in the modeled and observed tide could overwhelm the IS4DVAR procedure (note that, uncorrected, altimetry observations include the instantaneous signal of sea level variability associated with the local tide), we took the following approach. A 1-year long ROMS simulation without assimilation, forced by observed tidal harmonics (Egbert et al., 2002) on the model perimeter, was run to compute the model's tidal harmonic sea level response. Next, the default corrections provided in the Radar Altimeter Database System (RADS; <http://rads.tudelft.nl>) were applied to de-tide the along-track altimeter data. Finally, the ROMS model tide harmonics were interpolated to the altimeter ground-track positions and used to introduce ROMS tidal variability to the altimeter data prior to assimilation. Thus the observations are adjusted to include the modeled tide. By this approach, SSH variability associated with tides is the same in the model and adjusted data, and therefore will not enter into the IS4DVAR cost function; the model-data misfit in SSH is then dominated by processes other than the tides. It is these processes, such as the SSH variability associated due to mesoscale eddies in the Slope Sea and movements of the Gulf Stream and shelf-slope fronts that we wish to correct and project to mesoscale variability in the subsurface density and velocity field.

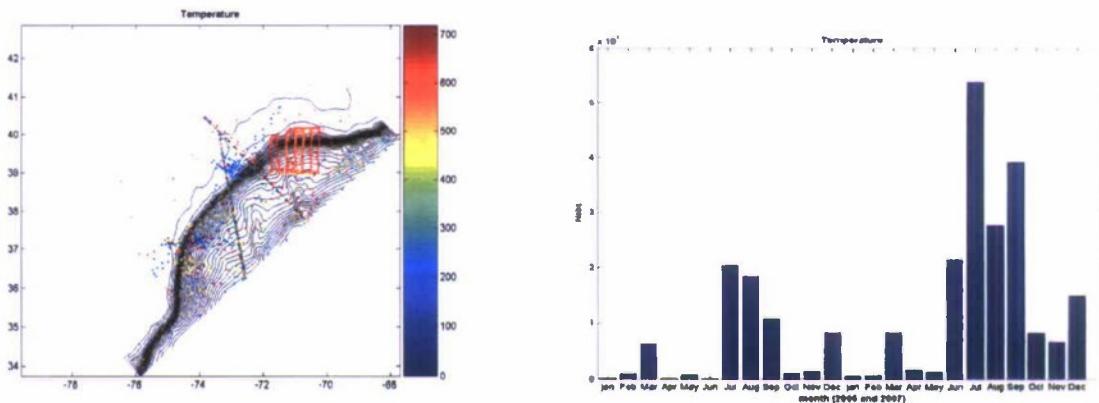
**Filtering of the barotropic mode:** The adjoint model can erroneously accommodate too much of the SSH model-data misfit in the barotropic mode, which sends fast gravity waves along the model perimeter. We therefore acknowledge the temporal correlation of the sub-tidal altimeter SSH data by repeating (duplicating) the altimeter SSH observations at  $t = -6$  hour,  $t = 0$  and  $t = +6$  hours, but with appropriate time lags in the added tide signal. These data cannot easily be matched by a barotropic wave. The benefit after assimilation of the altimeter data is illustrated in Fig 3 which compares the variance of the non-assimilative model versus that after assimilation. Figure 4 shows a correlation between model and observed SSH

The impact of the assimilation of SSH and SST in forecasting non-assimilated temperature profiles is illustrated in Fig 5, where all the temperature profiles for 2006 where ranked by water depth and plotted along with the ESPreSSO model predictions.



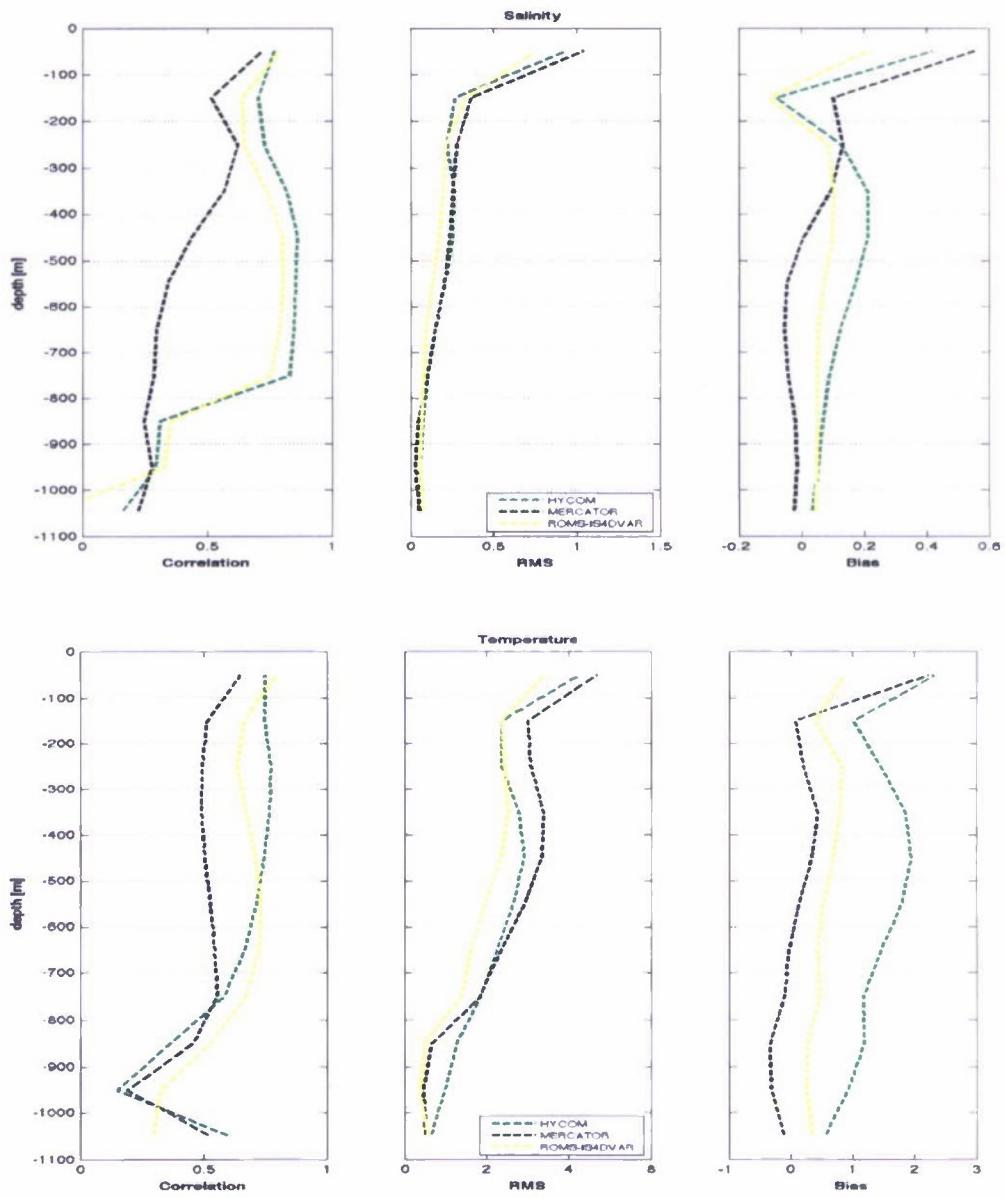
*Figure 4: ESPreSSO hindcast of observed temperatures for different water depths during 2006. Satellite information (SSH and SST) was only assimilated.*

In the late stages of the project we decided to extend our validation to include observations from year 2007 (mainly from the pioneer array) to obtain a more robust estimate of the skill of the model. Figure 5 shows the spatial and temporal distribution of the temperature observations during the two years period considered. Salinity exhibits a similar distribution.



*Figure 5: Spatial and temporal distribution of temperature observations from CTD or glider considered in the validation. In the left panel the geographical location is color-coded by the day (after Jan 1<sup>st</sup> 2006) when the observation was taken. The blue H-shaped observational network depicts the SW06 experiment and the orange matrix depicts the pioneer array. Most of the observations considered were taken during summer time and in the shelf break.*

These observations were binned by depth every hundred meters and then the model estimate in space and time was determined via linear interpolation of daily average files. Then within each depth-bin the correlation, RMS error and bias were estimated as an objective measure of the model skill. Figure 6 shows this validation for salinity and temperature, and as a reference we include the same validation for two other data-assimilative models (HYCOM and MERCATOR).



*Figure 5: Skill of ESPreSSO in nowcasting not-assimilated profiles of salinity and temperature during 2006 and 2007. For comparison the skill of two other data-assimilative models (HYCOM and MERCATOR) is included.*

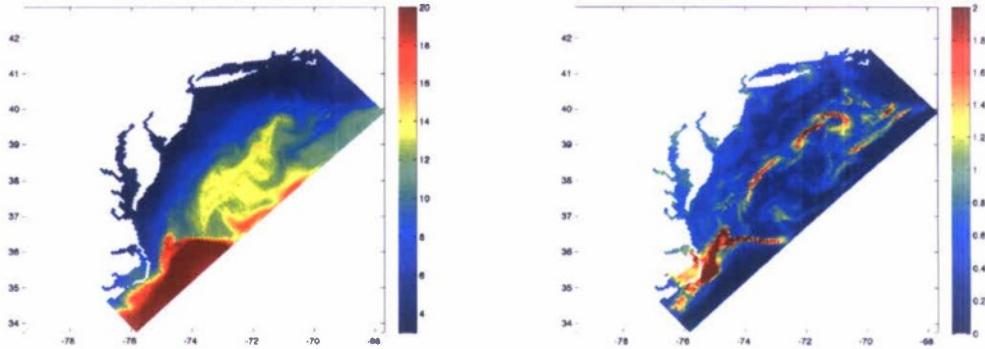
The current skill of ESPreSSO is superior to that of MERCATOR and at least comparable to that of HYCOM in terms of correlation and RMS error but significantly better in terms of bias. Detailed analysis of our data-assimilation system indicated that the strong seasonality in the meso and submesoscale variability has to be taken into account in the background error covariance matrix used in the assimilation. While this was beyond the scope of this project, this potential improvement is being tested in an updated version of ESPreSSO.

**Task 3: Development of an adjoint-based ensemble prediction system:** We developed an ensemble forecast system to determine whether there are persistent regions where uncertainty in the forecasted model state impacts predictive skill. This is an important addition in a real-time system because we don't know precisely the atmospheric forcing in advance. The ensemble members are produced by considering the effect of different atmospheric stochastic forcing realizations estimated from observations.

Synthetic timeseries of expected forcing  $F_s$  can be generated using at least two methodologies: (i) they can be generated as a linear combination of the dominant empirical orthogonal functions of the historical atmospheric analyses whose amplitude coefficients are modeled with the appropriate autoregressive process, and (ii) by melding bootstrapped timeseries of length typical of the atmospheric decorrelation time based on historical atmospheric analyses (details of this technique are presented in Zavala-Garay, 2008). Both techniques preserve key aspects of the atmospheric weather such as spatial coherence, seasonality in spatial distribution, climatological strength, and decorrelation times. In general the second technique is more robust in that it can model synthetic timeseries having a temporal behavior more complex than a simple autoregressive model, such as those associated with eastward propagating convective process (ZG08). In the ESPreSSO system we use the second technique to generate timeseries whose statistics mirrors that of the historical data. The synthetic timeseries can be then combined with the atmospheric forecast to produce an ensemble of forcings  $Fe$  using the formula

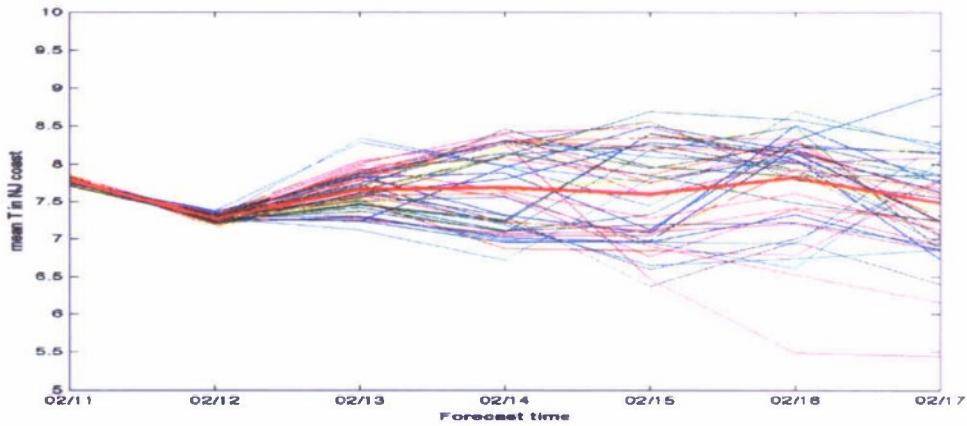
$$Fe = c * F + (1-c) * F_s, \quad (1)$$

where  $c = \exp(-t/\tau)$  is a time-dependent weight that reflects the credibility of the forecast (typically of the order of  $\tau = 5$  days). Note that each ensemble  $Fe$  tends to the forecasted value  $F$  for short periods of time ( $t < \tau$ , say 1 day) but its statistics tends to that of the historical data for long period of time ( $t > \tau$ , say 2 weeks). This ensemble of expected atmospheric forcings can then be used to produce an ensemble of ocean forecasts whose associated probabilities can then be used in decision making strategies. For instance, Figure 6 shows the 6-days-lead forecasted SST for Feb 17 2010 (initialized Feb 10, 2010) and the associated uncertainty given 50 atmospheric realizations produced formula (1). It can be seen that for this particular week the major uncertainty in the forecasted SST occurs in the frontal edge of the Gulf Stream and some parts of the shelf break.



**Figure 6:** ESPreSSO forced SST (left) and its associated error (right) six days in advance for Feb 17, 2010.

This methodology allows us to identify regions where the operational deterministic forecast is prone to large errors due to the inherent chaotic nature of the atmospheric forcing. A novel component of this project was to use of the adjoint model to estimate the expected variability of selected indices of the model forecast due to the action of atmospheric weather. The methodology is based on the use of a variant of the adjoint sensitivity analysis and it allows estimation of the uncertainty in the predicted indices using a few integrations of the adjoint model, in remarkable contrast with the direct method based on a large number of integrations of the forward model which in an operational framework is difficult to perform. The indices are chosen based on the application of interest such as a metric quantifying interchange of tracers across the shelf-break, the strength of shelf-break fronts (which coincide with regions of high acoustic uncertainty), the dispersion of a passive tracer (e.g., a contaminant), stratification in a selected area, or expected uncertainty in the predicted patterns of currents. As an example Figure shows an ensemble of forecast six days in advance for the temperature in front of the New Jersey coast (the index is the spatial mean temperature at each day).



**Figure 7.** Six-days ensemble forecasted SST on Feb 10 2010. The thick red line is the ensemble mean and the thin lines are each one of the ensemble members.

The addition of this final component will prototype a fully operational dynamical/statistical forecast system in a coastal transition zone that are of increasing interest to Navy operational requirements.

## IMPACT/APPLICATIONS

The prototype data assimilation system is now operating with real time data streams, which allow us to deliver real time forecast for the MAB region. The realtime forecasts are published at <http://www.myroms.org/espresso/>.

## TRANSITIONS

We have now produced a first version of reanalysis for 2006 and 2007 and this reanalysis has been made available to the community through an OPeNDAP server at <http://www.myroms.org/espresso/>. While this is beyond the scopes of this project, currently we are evaluating an improved hindcast system that will consider seasonal variations of background error covariances which we expect to highly improve the version currently available via OPeNDAP. Once evaluation of this new version is completed, the scientific findings will be published in scientific journals.

## REFERENCES

Ducet, N., P. Le Traon, and G. Reverdin, Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2, *J. Geophys. Res.*, 105, 19,477-19,498, 2000.

Egbert, G. D., and S. Y. Erofeeva, Efficient Inverse Modeling of Barotropic Ocean Tides, *Journal of Atmospheric and Oceanic Technology*, 19, 183-204, 2002.

Wilkin, J. L., M. M. Bowen, and W. J. Emery, Mapping mesoscale currents by optimal interpolation of satellite radiometer and altimeter data. *Ocean Dynamics*, 52, 95-103, 2002.

Zavala-Garay J., C. Zhang, A. M. Moore, A. Wittenberg, M. Harrison, A. Rosati, and J. Vialard, 2008: Sensitivity of hybrid ENSO models to uncoupled atmospheric variability. *J. Climate*, 21, 3704-3721.

Zavala-Garay J., J. L. Wilkin, and H. Arango, 2010: Predictability of mesoscale variability in the East Australia Current given strong-constraint data assimilation. *Journal of Physical Oceanography*. Accepted after minor revisions. Manuscript available at <ftp://marine.rutgers.edu/pub/jzavala/ms/EAC.pdf>